

## Studi Empiris Model BERT dan DistilBERT: Analisis Sentimen pada Pemilihan Presiden Indonesia

Mahira Putri<sup>1✉</sup>, Taufik Edy Sutanto<sup>2</sup>, Suma Inna<sup>3</sup>

[mahira.putri20@mhs.uinjkt.ac.id](mailto:mahira.putri20@mhs.uinjkt.ac.id)

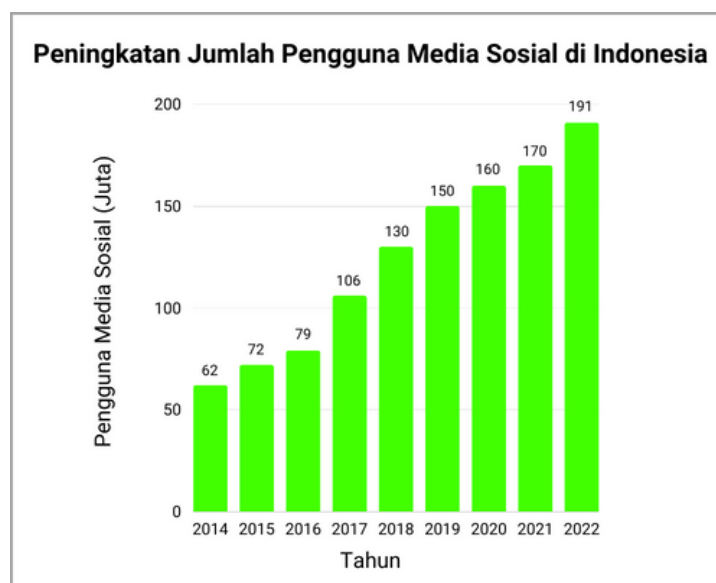
<sup>1,2,3</sup>UIN Syarif Hidayatullah Jakarta

Informasi Artikel	Abstrak
Diterima : 15 Okt 2023 Direview : 24 Okt 2023 Disetujui : 30 Okt 2023	Peningkatan jumlah pengguna media sosial di Indonesia sejak tahun 2014 menyebabkan data yang dihasilkan semakin besar dan kompleks, sehingga komputasi yang diperlukan untuk mengolahnya juga semakin besar. Untuk melakukan komputasi pada data yang besar diperlukan model yang kompatibel, efektif, dan efisien. Penelitian ini adalah kajian numerik dari dua model terbaik <i>Deep Learning</i> saat paper ini ditulis, yaitu BERT dan DistilBERT pada kasus analisis sentimen menggunakan ratusan ribu tweet terkait pemilihan presiden Indonesia tahun 2014 dan 2019. Analisis yang dilakukan meliputi waktu eksekusi dan konsumsi memori. Pada model dengan nilai <i>hyperparameter</i> optimal, tercatat bahwa DistilBERT melakukan proses pelatihan dan prediksi 84% lebih cepat dengan penggunaan memori GPU 79% lebih efisien dengan nilai akurasi tidak terpaut jauh, yaitu 0.89 dan 0.85 untuk BERT dan DistilBERT. Hasil kajian ini dapat digunakan untuk memperkirakan besarnya sumberdaya komputasi atau biaya yang dibutuhkan ketika menggunakan model BERT atau DistilBERT pada data yang besar.
<b>Kata Kunci</b>	
Analisis Sentimen, BERT, DistilBERT, Komputasi, Memori	

Keywords	Abstrak
<i>BERT, Computing, DistilBERT, Memory, Sentiment Analysis</i>	<i>The rapid growth in Indonesian social media users since 2014 has generated larger and more complex datasets that required substantial computational resources. Efficient, compatible models are crucial for processing such data. This study numerically compares BERT and DistilBERT, two leading Deep Learning models for sentiment analysis using hundreds of thousands of tweets related to the Indonesian presidential election in 2014 and 2019. Execution time and memory usage are examined. Optimal DistilBERT performs training and predictions up to 84% faster and uses GPU memory 79% more efficiently while maintaining comparable accuracy (0.89 for BERT and 0.85 for DistilBERT). These findings help estimate computational resources and costs when employing BERT or DistilBERT models on large data.</i>

## A. Pendahuluan

Aktivitas yang serba digital telah membuat jumlah pengguna media sosial di Indonesia terus mengalami peningkatan sejak 2014, hal ini ditunjukkan dengan *'Indonesian Digital Report 2022'* yang diterbitkan oleh Hootsuite (*We are Social*) [1] yang dapat dilihat pada Gambar 1. Dampak dari adanya peningkatan jumlah pengguna media sosial adalah data yang dihasilkan dari berbagai *platform* media sosial pun semakin besar dan kompleks, sehingga dibutuhkan komputasi yang lebih besar untuk mengolah data tersebut. Keterbatasan memori *bandwidth* pada *Dynamic Random-Access Memory* (DRAM) kerap kali menjadi hambatan. Sebagai contoh, walaupun telah menggunakan *mini-batch* 32 pada tipikal GPU berperforma tinggi, ResNet-50 menunjukkan bahwa pelatihan masih membutuhkan DRAM lokal lebih dari 7.5 GB [2].



**Gambar 1.** Peningkatan jumlah pengguna berbagai media sosial di Indonesia

Dalam pengaplikasiannya, data dan kebutuhan akan memori yang semakin membesar kerap kali menjadi tantangan pada proses komputasi, hal-hal tersebut kemudian menjadi alasan diperlukannya model yang lebih efisien namun tetap efektif. Dalam konteks analisis sentimen pada data teks yang besar, penggunaan model *Transformer* [3] menjadi salah satu solusi pada kendala yang dimiliki model-model berbasis *Recurrent Neural Network* (RNN), seperti proses komputasi yang lambat dan tidak dapat diparalelkan karena harus memproses urutan, hingga kesulitan dalam memodelkan ketergantungan jangka panjang. BERT [4] (*Bidirectional Encoder Representations from Transformers*) yang dikembangkan oleh Google AI merupakan salah satu model yang memanfaatkan arsitektur *Transformer*. DistilBERT yang merupakan versi destilasi BERT yang lebih sederhana dengan tetap mempertahankan sebagian besar kemampuan BERT [5].

Penelitian ini meneliti performa model DistilBERT yang diyakini memiliki kebutuhan komputasi yang lebih kecil dibanding BERT karena memiliki arsitektur yang lebih ringkas. Karena arsitektur yang lebih sederhana, penggunaan memori yang diperlukan DistilBERT juga cenderung lebih efisien karena jumlah parameter yang lebih sedikit. Namun demikian, perlu kajian seberapa besar efisiensi dapat

diperoleh dari pendekatan ini. Kemudian seberapa akurat model sederhana tersebut dalam melakukan prediksi.

Penelitian ini menyajikan pendekatan yang baru, karena belum ada penelitian yang melakukan studi empiris terhadap model BERT dan DistilBERT dalam konteks analisis sentimen dalam bahasa Indonesia. Selain itu, penelitian ini menggunakan kasus unik data Twitter pada pemilihan presiden (pilpres) Indonesia 2014 dan 2019. Tiga buah metrik *precision*, *recall*, dan *F1-score* digunakan untuk mengukur kinerja model dengan memperhatikan konsumsi memori dan komputasi melalui waktu yang diperlukan oleh masing-masing model untuk melakukan proses latih dan prediksi.

Devlin et al. [4] memperkenalkan BERT sebagai model representasi bahasa baru yang dirancang untuk dapat melakukan pelatihan representasi dua arah dari teks yang tidak berlabel. GLUE merupakan kumpulan alat yang digunakan untuk mengevaluasi sejauh mana kinerja yang dilakukan model-model pada berbagai tugas pemahaman bahasa alami yang ada [6], sedangkan MultiNLI adalah dataset yang dibuat untuk memahami kalimat dalam konteks pengembangan dan evaluasi model [7]. BERT dilaporkan dapat meningkatkan skor GLUE menjadi 80.5%, dan akurasi MultiNLI menjadi 86.7%. Kemudian, Sanh et al. [5] mengusulkan DistilBERT, model representasi bahasa dengan kebutuhan lebih kecil, yang dapat disesuaikan dengan berbagai tugas dalam model yang besar. Dengan memanfaatkan pengetahuan BERT yang telah disederhanakan selama tahap pra-pelatihan, DistilBERT menunjukkan kemampuannya untuk mengurangi ukuran BERT hingga 40%, memberikan waktu komputasi hingga 60% lebih cepat, dengan tetap mempertahankan kemampuan BERT sebesar 97%. Said dan Manik [8] memprediksi pemenang pilpres Indonesia 2019 pada data Twitter menggunakan BERT dan RoBERTa, dan menghasilkan akurasi mencapai 98.02%. Selain itu, BERT mengklasifikasi multi-label dengan akurasi tertinggi mencapai 98.66%.

Batra et al. [9] mengeksplorasi BERT dan model-model berbasis BERT untuk menganalisis komentar pada GitHub, Jira, dan unggahan *Stack Overflow*. Terdapat tiga strategi yang dilakukan, strategi pertama adalah menyesuaikan model BERT yang telah dilatih dan disimpulkan bahwa *F1-score* paling rendah dimiliki oleh model ALBERT pada *dataset Stack Overflow* sebesar 0,67, sementara itu RoBERTa memberikan hasil lebih baik pada *dataset GitHub* dan *Jira* dengan *F1-score* masing-masing adalah 0,91 dan 0,83. Strategi kedua adalah mengembangkan model *ensemble* dari BERT, pendekatan *ensemble* menjadi motivasi untuk mendapatkan model yang lebih kuat dengan kesimpulan yang sedikit bervariasi dari model RoBERTa, ALBERT, dan BERT-base, di mana *F1-score* untuk *dataset GitHub* meningkat menjadi 0,92. Strategi terakhir adalah menggunakan model BERT yang dikompresi (DistilBERT), *F1-score* pada *Stack Overflow* mengalami peningkatan menjadi 0,88 dan pada *Jira* menjadi 0,84.

Fajri et al. [10] menganalisis perbandingan BERT dan DistilBERT pada *dataset Twitter* mengenai Covid-19 dan mengklasifikasikannya menjadi *Positive*, *Extremely Positive*, *Negative*, *Extremely Negative*, dan *Neutral*. BERT menghasilkan akurasi sebesar 87%, *precision* dan *recall* tertinggi sebesar 91% ditunjukkan berturut-turut oleh sentimen *Neutral* dan *Extremely Negative*, *f1-Score* tertinggi sebesar 89% diperoleh oleh *Neutral* dan *Extremely Positive*. DistilBERT menghasilkan akurasi sebesar 97%, *precision* tertinggi sebesar 99% didapat oleh sentimen *Neutral*, *recall*

dan *f1-score* tertinggi sebesar 99% ditampilkan pada sentimen *Extremely Negative* dan *Extremely Positive*.

Cortiz [11] menguji beberapa model pada dataset GoEmotion, yang digunakan untuk mengidentifikasi emosi dengan tingkat granularitas tinggi dan diannotasikan secara manual. Dengan membandingkan kinerja model-model berbasis transformer seperti BERT (sebagai model dasar), DistilBERT, RoBERTa, XLNet, dan ELECTRA, menunjukkan F1-score sebesar 46% untuk BERT, 48% untuk DistilBERT dan XLNet, 49% untuk RoBERTa, dan ELECTRA memiliki F1-score terendah yaitu 33%. Disamping itu, Adoma et al. [12] menggunakan dataset ISEAR saat melakukan analisis perbandingan model BERT, RoBERTa, DistilBERT, dan XLNet pada tugas pengenalan emosi berbasis teks. Hasil pada *confusion matrix* menampilkan akurasi model RoBERTa, XLNet, BERT, dan DistilBERT berturut-turut sebesar 0.7431, 0.7299, 0.7009, 0.6693.

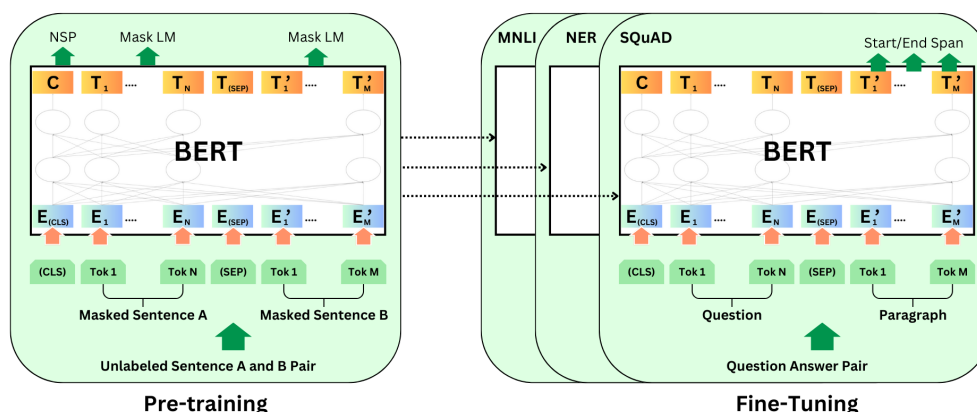
Hal yang menjadi pembeda dari kajian ini dengan kajian-kajian yang telah dijelaskan sebelumnya adalah analisis performa model tidak hanya melalui akurasi, namun juga melalui konsumsi memori dan waktu yang diperlukan untuk melakukan proses komputasi.

## B. Metode Penelitian

Bagian ini menjelaskan dengan singkat model Transformer yang digunakan dalam kajian ini, yaitu BERT dan DistilBERT. Hal yang dibahas mencakup penjelasan mengenai layer dan jumlah parameter, serta alasan penggunaan kedua model dalam penelitian ini.

### 1. BERT

BERT [4] (*Bidirectional Encoder Representations from Transformers*) digunakan sebagai model dasar (*baseline*) pada kajian ini. Gambar 2 memberikan gambaran mengenai prosedur dan arsitektur pada BERT. Saat *pre-training*, model dilatih dengan tugas *pre-training* yang berbeda pada data yang tidak memiliki label. Dilanjut dengan proses *fine-tuning*, yaitu proses inialisasi BERT dengan parameter *pre-training*. Parameter sebelumnya akan dilatih kembali (*fine-tuned*) menggunakan data yang telah dilabeli dari tugas turunan (*downstream tasks*).



**Gambar 2.** Prosedur *Pre-Training* dan *Fine-Tuning* pada Arsitektur BERT [4]

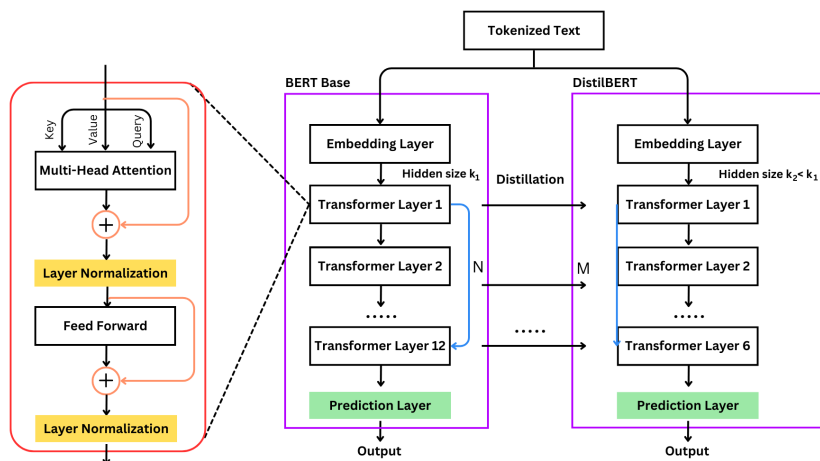
*Self-attention* atau juga disebut *intra-attention* adalah sebuah mekanisme yang bekerja dengan memusatkan perhatian pada data berurutan [3]. *Self-attention*

memungkinkan model untuk dapat memberikan bobot yang berbeda pada setiap elemen dalam urutan yang tergantung pada hubungan antar elemen.

BERT Base memiliki 12 *layer*, 768 *hidden size*, 12 *self-attention head*, dan total parameter sebanyak 110 juta. Sedangkan BERT Large memiliki 24 *layer*, 1024 *hidden size*, 16 *self-attention head*, dan total parameter sebanyak 340 juta. Hal ini menjadi salah satu aspek yang membuat BERT memiliki kemampuan yang lebih baik dalam memahami kosakata, sintaks, dan hubungan semantik dalam teks secara lebih mendalam.

## 2. DistilBERT

DistilBERT adalah varian dari model BERT yang telah disederhanakan dengan mengurangi kompleksitas dan ukuran arsitekturnya, dengan tetap mempertahankan 97% kemampuan pemahaman bahasa yang dimiliki oleh BERT [5]. DistilBERT hanya memiliki 6 *layer* (50% dari BERT) yang dalam proses penyederhanaannya, terdapat proses penggabungan informasi dari beberapa *layer* pada BERT. Bobot (*weights*) dari *layer* BERT yang lebih dalam, akan disesuaikan dengan *layer* yang lebih dangkal pada DistilBERT kemudian ditransfer ke dalamnya, sehingga DistilBERT hanya memiliki jumlah parameter sebanyak 66 juta yang selanjutnya dapat dilihat pada Gambar 3.



**Gambar 3.** Komponen dan Arsitektur DistilBERT [13]

Langkah penting dalam pengoptimalan DistilBERT adalah dengan mengambil satu dari dua lapisan dalam inialisasi dan dengan memanfaatkan kemiripan struktural antara BERT dan DistilBERT, Hal ini yang menjadikan DistilBERT lebih sederhana sehingga dapat meminimalkan ukurannya dan mempercepat proses komputasinya.

## C. Hasil dan Pembahasan

Bagian ini akan membahas rangkaian eksperimen yang dilakukan, mencakup data, IndoBERT, dan *performance analysis*. Pada bagian data, akan dibahas mengenai data yang digunakan, termasuk sumber data dan prapemrosesan. Selanjutnya pembahasan mengenai penggunaan IndoBERT yang dikhususkan untuk dataset bahasa Indonesia. Terakhir adalah pembahasan mengenai analisis performa dari model.

## Data

Sebanyak 100.000 baris data untuk masing-masing tahun 2014 dan 2019 dengan topik Pilpres di Indonesia, dikumpulkan secara legal dan etis serta menghormati privasi pengguna platform media sosial Twitter. Data terdiri dari beberapa variabel seperti kapan *tweet* dibuat, nama pengguna dan jumlah *retweet*, namun dalam penelitian ini hanya menggunakan variabel "*tweet*". Data teks kemudian diklasifikasikan menjadi: positif, negatif, dan netral. Namun, untuk analisa selanjutnya hanya kategori positif dan negatif yang ditelaah lebih lanjut.

Karakter khusus, spasi ganda, tag pengguna (*username*), tautan, *retweet*, *hashtag*, dan elemen lain yang dapat mempengaruhi kinerja secara negatif dihapus, begitu pula dengan *stopwords*, *NaN*, dan data duplikat. Selain itu, teks dalam variabel "*tweet*" diformat ke dalam huruf kecil (*lowercase*) kemudian ditempatkan dalam variabel baru yaitu "*cleaned\_tweet*". Setelah itu, dilakukan penggantian pada label sentimen, yaitu negatif=0 dan positif=1. Data yang telah melalui tahap pra-pemrosesan kemudian dibagi menjadi dua kelompok, yaitu 90% untuk data pelatihan dan 10% untuk data pengujian. Kalimat terpanjang dalam "*cleaned\_tweet*" memiliki panjang 49 kata, sehingga panjang maksimum kalimat yang ditetapkan untuk tokenizer adalah 51.

## IndoBERT

IndoBERT [14] merupakan model *pre-trained* yang mengadaptasi arsitektur BERT khusus untuk bahasa Indonesia. Proses *pre-trained* telah diterapkan pada IndoBERT dengan dataset bahasa Indonesia yang berukuran besar, sehingga model ini memiliki pemahaman yang lebih baik terhadap bahasa tersebut. Selanjutnya, IndoBERT dapat di *fine-tuned* pada permasalahan NLP berbahasa Indonesia, memungkinkannya mampu memahami struktur bahasa, kosakata, dan konteks bahasa Indonesia lebih spesifik pada penelitian ini.

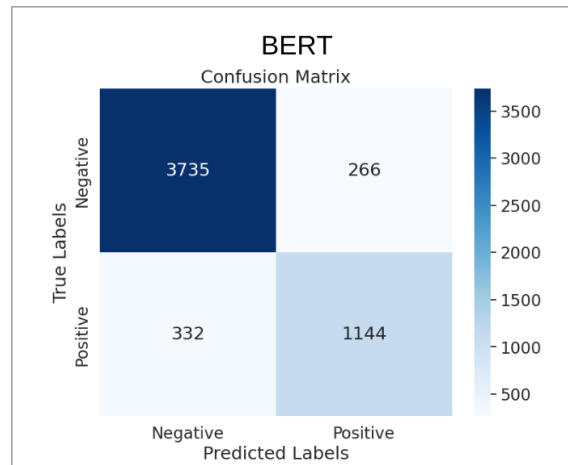
## Analisis Performa

Secara umum, hyperparameter yang digunakan pada penelitian ini telah diatur sebagian besar mengikuti pengaturan yang ditentukan oleh Devlin et al. [4] Hal yang disesuaikan hanyalah *learning rate* ( $2e^{-5}$ ) dan *batch size* (32). Kedua model dilatih dengan *environment* GPU NVIDIA V100.

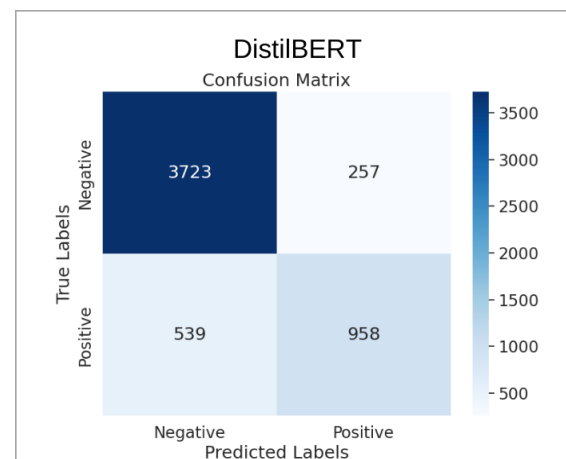
Gambar 4 dan Gambar 5 menampilkan *confussion matrix* dari model BERT dan DistilBERT. Pada Gambar 4, terlihat bahwa BERT dapat memprediksi dengan benar label negatif sebanyak 3735 dari 4001 label dan label positif sebanyak 1144 dari 1476 label. Sedangkan pada Gambar 5, terlihat bahwa DistilBERT dapat memprediksi dengan benar label negatif sebanyak 3723 dari 3980 label dan label positif sebanyak 958 dari 1497 label. *Confussion matrix* yang digunakan mengacu pada *macro average* yaitu dengan menghitung *matrix* dari setiap *confussion matrix* kemudian menghitung rata-rata skor *matrix* tersebut.

Secara keseluruhan BERT menghasilkan akurasi, presisi, *recall*, dan *f1-score* yang lebih besar dari DistilBERT. Akurasi yang dihasilkan BERT sebesar 0.89 dan DistilBERT sebesar 0.85. Waktu eksekusi pada proses *training* dan *validation* pada BERT juga cenderung lebih lama. Hal ini sangat menarik, karena DistilBERT hanya membutuhkan waktu selama 2 menit 28 detik untuk *training*, dimana BERT membutuhkan waktu selama 13 menit 21 detik lebih lama. Terkait penggunaan

memori, DistilBERT mengonsumsi memori RAM 134.46 MB lebih besar dari BERT. Sedangkan penggunaan memori GPU [15] menunjukkan BERT menggunakan memori GPU lebih besar.



**Gambar 4.** Confusion Matrix Model BERT



**Gambar 5.** Confusion Matrix Model DistilBERT

**Tabel 1.** Perbandingan Akurasi, Presisi, Recall, F1-score, Waktu Eksekusi, dan Konsumsi Memori pada BERT dan DistilBERT

Model	Akurasi	Presisi	Recall	F1-score	Waktu Eksekusi		Konsumsi Memori	
					Training	Validation	RAM	GPU
BERT	0.89	0.86	0.85	0.86	0:15:49	0:00:30	174.89 MB	15.14 GB
DistilBERT	0.85	0.83	0.79	0.80	0:02:28	0:00:04	309.35 MB	3.17 GB

Langkah-langkah yang terjadi saat menghitung penggunaan memori adalah pemuatan model, *forward pass*, *backward pass*, langkah pengoptimalan, dan jalankan iterasi selanjutnya. Berikut rumus konsumsi memori maksimal saat pelatihan:

$$\text{konsumsi memori maksimal} = m + f * \text{batch size} * b + d * g + o * m$$

Dengan :

$m$  = memori model,

$f$  = banyaknya memori yang digunakan forward pass untuk  $batch\_size = 1$ ,  
 $g = m$ , adalah banyaknya memori untuk gradien,  
 $d = 1$  jika pelatihan menggunakan 1 GPU, dan 2 jika menggunakan  $>1$  GPU,  
 $o$  = banyaknya momen yang disimpan oleh pengoptimal,  
 $b = 0.5$  untuk pelatihan dengan presisi campuran, 1 untuk pelatihan dengan presisi penuh.

Sehingga diperoleh GPU *memory usage* BERT adalah 15.14 GB dan DistilBERT 3.17GB.

#### D. Simpulan

Paper ini melakukan studi empiris pada model transformer BERT dan versi destilasinya (DistilBERT) pada studi kasus analisis sentimen data Twitter pilpres Indonesia 2014 dan 2019. Lamanya waktu eksekusi dan banyaknya konsumsi memori mengkonfirmasi efektivitas dan efisiensi DistilBERT dibandingkan dengan BERT dalam menganalisis sentimen data Twitter pilpres tersebut. Analisis yang dilakukan meliputi aspek waktu eksekusi dan konsumsi memori, tercatat bahwa DistilBERT dapat melakukan proses pelatihan dan prediksi hingga 84% lebih cepat dengan penggunaan memori GPU 79% lebih efisien dengan nilai akurasi yang tidak terpaut jauh, yaitu 0.89 dan 0.85 untuk BERT dan DistilBERT.

#### E. Referensi

- [1] A. D. Riyanto, "Hootsuite (We are Social): Indonesian Digital Report 2022." [Daring]. Tersedia pada: <https://andi.link/hootsuite-we-are-social-indonesian-digital-report-2022/>
- [2] J. Hanlon, "Why is So Much Memory Needed for Deep Neural Networks?," GRAPHCORE. Diakses: 7 Oktober 2023. [Daring]. Tersedia pada: <https://www.graphcore.ai/posts/why-is-so-much-memory-needed-for-deep-neural-networks>
- [3] A. Vaswani *dkk.*, "Attention is All you Need," dalam *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, dan R. Garnett, Ed., Curran Associates, Inc., 2017. [Daring]. Tersedia pada: [https://proceedings.neurips.cc/paper\\_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf)
- [4] J. Devlin, M.-W. Chang, K. Lee, dan K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," 2018, doi: 10.48550/ARXIV.1810.04805.
- [5] V. Sanh, L. Debut, J. Chaumond, dan T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," 2019, doi: 10.48550/ARXIV.1910.01108.
- [6] A. Wang, A. Singh, J. Michael, F. Hill, O. Levy, dan S. Bowman, "GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding," dalam *OpenReview.net*, New Orleans, Louisiana, United States, Des 2018. [Daring]. Tersedia pada: [https://openreview.net/forum?id=r\]4km2R5t7](https://openreview.net/forum?id=r]4km2R5t7)
- [7] A. Williams, N. Nangia, dan S. Bowman, "A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference," dalam *Proceedings of the 2018 Conference of the North American Chapter of the Association for*

- Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, New Orleans, Louisiana: Association for Computational Linguistics, 2018, hlm. 1112–1122. doi: 10.18653/v1/N18-1101.
- [8] F. Said dan L. P. Manik, “Aspect-Based Sentiment Analysis on Indonesian Presidential Election Using Deep Learning,” *Paradig. - J. Komput. Dan Inform.*, vol. 24, no. 2, hlm. 160–167, Sep 2022, doi: 10.31294/paradigma.v24i2.1415.
- [9] H. Batra, N. S. Punj, S. K. Sonbhadra, dan S. Agarwal, “BERT-Based Sentiment Analysis: A Software Engineering Perspective,” dalam *Database and Expert Systems Applications*, vol. 12923, C. Strauss, G. Kotsis, A. M. Tjoa, dan I. Khalil, Ed., dalam *Lecture Notes in Computer Science*, vol. 12923. , Cham: Springer International Publishing, 2021, hlm. 138–148. doi: 10.1007/978-3-030-86472-9\_13.
- [10] F. Fajri, B. Tutuko, dan S. Sukemi, “Membandingkan Nilai Akurasi BERT dan DistilBERT pada Dataset Twitter,” *JUSIFOJ. Sist. Inf.*, vol. 8, no. 2, hlm. 71–80, Des 2022, doi: 10.19109/jusifo.v8i2.13885.
- [11] D. Cortiz, “Exploring Transformers in Emotion Recognition: a comparison of BERT, DistilBERT, RoBERTa, XLNet and ELECTRA,” 2021, doi: 10.48550/ARXIV.2104.02041.
- [12] A. F. Adoma, N.-M. Henry, dan W. Chen, “Comparative Analyses of Bert, Roberta, Distilbert, and Xlnet for Text-Based Emotion Recognition,” dalam *2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, Chengdu, China: IEEE, Des 2020, hlm. 117–121. doi: 10.1109/ICCWAMTIP51612.2020.9317379.
- [13] H. Adel *dkk.*, “Improving Crisis Events Detection Using DistilBERT with Hunger Games Search Algorithm,” *Mathematics*, vol. 10, no. 3, hlm. 447, Jan 2022, doi: 10.3390/math10030447.
- [14] F. Koto, A. Rahimi, J. H. Lau, dan T. Baldwin, “IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP,” 2020, doi: 10.48550/ARXIV.2011.00677.
- [15] J. Stern, “A comprehensive guide to memory usage in PyTorch,” Medium. [Daring]. Tersedia pada: <https://medium.com/deep-learning-for-protein-design/a-comprehensive-guide-to-memory-usage-in-pytorch-b9b7c78031d3>